Students' Use of Alternative Scientific Causal Models in Explanations about Evolutionary Change: Selective vs. Stochastic Reasoning[1]

Elizabeth P. Beggrow[2]

Abstract

Non-adaptive, stochastic evolutionary causes, such as genetic drift, comprise an important element of biologists' explanatory models of evolutionary change, and yet science education research has focused almost exclusively on student ideas and misconceptions about natural selection. After instruction that includes stochastic causal factors (such as genetic drift), how do students construct evolutionary explanations? We used clinical interviews, open-response and multiple-choice instruments to investigate undergraduate students' non-adaptive reasoning (NAR) patterns. After instruction, we found NAR to be very uncommon in students' explanatory models of evolutionary change in both written assessments and clinical interviews. However, when NAR was used by students, it was conceptualized in an expert-like way; that is, non-adaptive and stochastic factors were modeled as alternatives to selection. Interestingly, non-adaptive reasoning was not found to be associated with greater understanding of natural selection in interviews or written assessments, or with fewer misconceptions of natural selection. Thus, NAR appears to be a distinct facet of evolutionary thinking. Greater attention to NAR in biology education is needed given how uncommonly students use it to explain evolutionary change.

[1] This is a version of an article published in *Evolution Education and Outreach*. The final publication is available at link.springer.com/article/10.1007/s12052-012-0432-z. Full Citation: Beggrow, E.P. and Nehm, R.H. 2012. Students' mental models of evolutionary causation: Natural selection and genetic drift. Evo Edu Outreach, 5(3), 429-444.

[2] Thanks to Ross Nehm, Minsu Ha, Meghan Rector Federer for their contributions to the research and data analyses, Judy Ridgway and Prof. Jon Harder for helping to access participants and the National Science Foundation REESE grant 099999 for financial support of portions of this study.

Introduction

A central goal of science education reform is to refocus teaching, learning, and assessment on core concepts or 'big ideas' (e.g., NRC, 2001a, 2001b; AAAS, 2011). One such big idea in the life sciences is biological evolution. Although evolution comprises the framework upon which the life sciences are structured (Dobzhansky, 1973), student learning of the idea and its associated causal concepts remains extremely problematic for students at all levels of the educational hierarchy (Bishop and Anderson, 1990; Nehm and Reilly, 2007; Gregory, 2009). In an effort to understand how to improve student learning about evolution, a major thrust of science education research has been to focus on students' alternative conceptions about evolution and develop pedagogies to initiate conceptual change (e.g., Demastes, Settlage and Good, 1995). Considerably less research has focused on how students reason about alternative *scientific* models seeking to explain the same phenomenon (e.g., of the possible causes, which most powerfully account for evolutionary change?). Indeed, within the field of science education, research has focused overwhelmingly on adaptive and selective thinking patterns. Research has yet to explore how non-adaptive factors are situated within students' conceptual ecologies of evolutionary causation.

While intense debate about the causal factors accounting for the structure of the natural world persists in many science domains, and reflects deeply complex philosophical and epistemological tensions (e.g., Conway Morris, 2003; Randall, 2011), we wish to engage with this topic at a level that intersects with important questions about the teaching of evolution. Within the field of evolutionary biology, deterministic and stochastic frameworks have been used to account for and explain (conceptually and methodologically) the processes that undergird evolutionary patterns in the living world, both past and present (Gould, 2002). While the body of work in this area is immense, we wish to focus attention on two concepts that represent core perspectives on evolutionary causation and have direct relevance to K-16 science educators: natural selection and genetic drift (Table 1).

[Insert Table 1 about here]

Natural selection is considered by evolutionary biologists to be the primary mechanism causing adaptive evolutionary change (Pigliucci and Müller, 2010). Adaptive evolutionary change occurs when the frequency of a trait (genetic or phenotypic feature) increases *because* it confers a survival or reproductive advantage to the 'individuals' (genes, bacteria, animals, etc.) possessing it (Dawkins, 1976; Darwin, 1859). Importantly, as noted above, natural selection is not the only cause of evolutionary changes in the living world. Alternative mechanisms, collectively known as 'non-adaptive' factors, include concepts such as genetic drift and developmental constraints, among others (Gould and Lewontin, 1979; Nickels, Nelson and Beard, 1996; Gould, 2002). Non-adaptive change (e.g., genetic drift) occurs when the frequency of a trait increases or decreases because of stochastic factors, regardless of whether the trait confers an advantage, disadvantage, or is neutral with respect to survival or reproduction (Freeman, 2005).

As noted by Orr (1998, p. 2099), evolutionary biologists have long been seeking ways to determine whether a phenotypic or genetic difference (e.g., short vs. long spines; dark vs. light coloration; presence or absence of an allele) was caused by natural selection, genetic drift, or combinations of the two factors. Biologists often conceptualize selection and drift as alternative or competing evolutionary explanations; selection is deterministic, and drift is stochastic (Sober 1984, p. 110). Methodologically, several approaches have been developed by evolutionary biologists to empirically test the causal contributions of these two processes (for details, see Orr, 1998). While many studies have modeled evolutionary change as being caused by either selection *or* drift, research has also looked at the collective contributions of both mechanisms to evolutionary change (i.e., selection *and* drift; Ackermann and Cheverud, 2004).

Mirroring broader scientific debates about the role of deterministic and stochastic processes in the natural world, the relative importance of selection and drift has been controversial within evolutionary biology for some time. Evolution by natural selection was introduced through the work of Darwin (1859) and Wallace (1871) and gained ground after the "eclipse of Darwinism" in the early twentieth century (Bowler, 1983). At this time, publications raised the question of the relative contributions of selection and random survival (e.g., Hagedoorn and Hagedoorn, 1921; Fisher, 1922). Fisher, for example, viewed deterministic processes as paramount and natural selection as the only important cause of evolutionary change (Provine, 1971). Genetic drift was introduced primarily through Wright's studies of population genetics in the early 1930s, but a lack of convincing experimental evidence led some evolutionary biologists, including Fisher, to reject drift as an important cause of evolutionary change (Savage, 1969; Provine, 1971).

These earlier debates were revived with Kimura's (1968) finding of a high rate of neutral mutation; such empirical findings provided support for stochastic causal contributors (e.g., genetic drift) to evolutionary change. Nevertheless, many evolutionary biologists remained strong subscribers of what some have termed an 'adaptationist approach' to evolution. For instance, Mayr argued that it is in fact the goal of the evolutionary biologist to first try 'to explain biological phenomena and processes as the product of natural selection' (1983, p. 326). Gould and Lewontin (1979) countered that Mayr's view is typical of what they caricatured as the 'Panglossian paradigm.' Gould and Lewontin refer to Voltaire's Dr. Pangloss of *Candide*, whose philosophy that ‘everything is for the best in this best of all possible worlds’ leads to numerous just-so explanatory stories, such as ‘our noses were made to carry spectacles, so we have spectacles’. The Panglossian paradigm serves as a metaphor for some biologists’ adaptationist approach to evolution, where the presence of traits is explained through the useful function of those traits. The Panglossian Paradigm demands deterministic explanations, thereby precluding any possibility of trait origins being a product of non-adaptive factors such as drift. Gould and Lewontin (1979, p. 587) warned of the dangers of such explanatory approaches: ‘One must not confuse the fact that a structure is used in some way…with the primary evolutionary reason for its existence and conformation.’ Likewise, Lynch (2007) viewed selection-centered views as reducing evolution to a form of ‘engineering’, which he considered not only unnecessary, but also misleading. Other scientists, in contrast, have taken a holistic approach to evolutionary change that recognizes the relative contributions of both causal factors (e.g., Parker and Maynard Smith, 1990; Orr, 1998). In short, evolutionary biologists have debated how evolutionary change should be framed, with some taking an adaptationist approach, and others taking a more pluralistic approach to evolutionary causation.

The competing roles of selection and drift in evolutionary causation have also been a subject of debate for philosophers of biology, but the importance of both concepts is *not* disputed (Sober, 1984). Discussions in the field of philosophy concern whether or not drift and selection are in fact two *distinct* concepts (e.g., Beatty, 1984; Sober, 1984; Matthen and Ariew, 2002; Millstein, 2002). Overall, in the philosophical and scientific communities, selection and drift are widely recognized by scientists and philosophers as two central evolutionary processes that should be considered when developing and testing explanations for trait differences between units (e.g., genomes, populations, species, etc.). Thus, the primary debate concerns the relative contributions of selection and drift.

Selection and drift in science education. Although stochastic processes are acknowledged as potential contributors to evolutionary change by biologists and philosophers (Gould, 2002; see above), life science curricula do not appear to reflect such scientific conceptualizations. For students to achieve expert-like competency in evolutionary reasoning, they must consider adaptive and non-adaptive processes as possible models of evolutionary causation. The National Association of Biology Teachers' position statement on teaching evolution (2011) mentions genetic drift as an example the various mechanisms behind the diversification and extinction of living organisms. Yet, in the *National Science Education Standards*, while natural selection is highlighted, non-adaptive concepts such as genetic drift are not mentioned at all (NRC, 1996). Given this paucity of attention in national and state *Standards*, it is likely that non-adaptive reasoning is scarce in most K-12 life sciences curricula, although no empirical studies appear to have investigated this issue.

Despite being of apparently peripheral focus in K-12 education, alternative mechanisms to natural selection are notably present in undergraduate and graduate biology programs and most college biology textbooks (e.g., Campbell, Reece, Urry, Cain, Wasserman, Minorsky and Jackson, 2008). Nevertheless, non-adaptive causal mechanisms are certainly not ubiquitous; indices of introductory biology textbooks indicate relatively low frequencies of topics related to 'genetic drift' (Campbell et al., 2008: 3 pages/1267 content pages; Futuyma, 1998: 32 pages/754 content pages; Freeman, 2005: 9 pages/1238 content pages). Furthermore, when present in textbooks, non-adaptive factors are usually insufficiently covered (Linhart, 1997). The same is true of laboratory manuals (Maret and Rissing, 1998). Considering the lack of emphasis on non-deterministic processes in the science standards (NRC, 1996) and the reliance on texts for teaching (Carpenter, Bullock and Potter, 2006; Cuseo, 2007), it is likely that most undergraduate science curricula over-emphasize natural selection at the expense of non-adaptive factors. It cannot be assumed that students receive appropriate instruction about non-adaptive causal mechanisms that would assist them in building pluralistic and expert-like conceptions of evolutionary causation. The possible consequences of a nearly exclusive focus on natural selection as the cause of evolutionary change are concerning and raise the question as to whether biology education is reifying an 'adaptationist' paradigm.

Although a number of studies have developed activities for teaching genetic drift in the classroom, or advocated for hands-on activities relating to genetic drift (e.g. Nickels, Nelson and Beard, 1996; Maret and Rissing, 1998; Staub, 2002), remarkably little work has been done to explore how (or if) students think about non-adaptive factors in evolutionary change. We know that practicing evolutionary biologists attribute evolutionary change to selection, stochastic processes (e.g., genetic drift) or a combination of these two factors, but how are students' models structured? This question motivated our explorations of biology students' non-adaptive and adaptive evolutionary reasoning patterns.

Theoretical Framework

It is well established that students bring a variety of intuitive ideas to school that are in conflict with normative scientific perspectives (e.g. Wandersee, Mintzes, & Novak, 1994). Students deal with these contradictions in a variety of ways: they may ignore the new information and continue using their previous frameworks; they may maintain both the new and old information in parallel, accessing each in specific contexts or situations; or they may construct a new conceptual framework that incorporates both the new and old knowledge (Fosnot, 1996). Students use their existing conceptual frameworks to process new experiences (e.g., assimilation) or, when the students' current frameworks are inadequate in allowing them to make sense of new experiences, they must reorganize and/or replace them with new concepts (e.g., accommodation) (Demastes, Good & Peebles, 1995; Posner, Strike, Hewson, & Gertzog, 1982; Sinatra, Brem & Evans, 2008). Students are actively and constantly engaged in processing information and ironing out contradictions through the process of equilibration (Piaget, 1964).

Students' alternative conceptions in science, particularly those regarding natural selection, are well documented (e.g., Bishop & Anderson, 1990; Settlage, 1994; Demastes, Settlage & Good, 1995; Demastes, Good & Peebles, 1995; Ferrari & Chi, 1998; Nehm and Reilly, 2007; Nehm and Schonfeld, 2007; Nehm, Kim, and Sheppard, 2009; Gregory, 2009). These alternative conceptions are both abundant and persistent, and efforts to target them through instruction have yielded varied results (e.g. Bishop & Anderson, 1990; Demastes, Settlage & Good, 1995; Demastes, Good & Peebles, 1995, 1996; Dagher & BouJaoude, 1997; Ferrari & Chi, 1998; Nehm and Reilly, 2007; Sinatra, Brem and Evans, 2008). In an effort to understand how to improve student learning and thinking about evolution, most of the research in evolution education has focused on students' alternative conceptions and how to change them (e.g., Demastes, Settlage & Good, 1995; Demastes, Good, & Peebles, 1995, 1996; Sinatra et al., 2008). Much less research has focused on how students incorporate alternative *scientific* models of evolutionary causation; that is, work has yet to explore how non-adaptive factors should be situated theoretically, or how non-adaptive factors such as genetic drift are incorporated into students' conceptual ecologies of evolution.

The development of evolutionary reasoning: theoretically situating non-adaptive causation. The National Research Council [NRC] (2001b) argues that an understanding of how students think and reason about domain-specific ideas should undergird the design of teaching, curriculum, and assessment. Such models should be based upon empirical evidence about how '…students represent knowledge and develop competence in the domain' (2001b, p. 3). Taken collectively, the aforementioned studies about student thinking and alternative concepts of evolution and natural selection contribute to the development of a model of how students think about and learn evolutionary concepts. Despite important early work in this area (cf. Catley, Lehrer, and Reiser, 2005), non-adaptive factors remain to be incorporated within cognitive models of evolutionary competence and corresponding learning progressions.

Novice-expert studies, common in many areas of science education, provide crucial insights into notions of 'competency' as well as what is meant by normative and accurate scientific understanding (NRC, 2001b). When comparing novices with experts, research demonstrates that differences between novice students and expert scientists lie in a variety of factors, including metacognition, organization and categorization of knowledge, and presuppositions surrounding knowledge (e.g., Vosniadou, 1999; Chi, Feltovich and Glaser, 1981; Authors 2011). Novice-expert studies have also compared knowledge representation and which concepts are used to build explanations to account for scientific phenomena (Keil and Wilson, 2000). Unfortunately, only one study has explored novice-expert reasoning patterns in evolution, with a focus almost exclusively on natural selection (Authors, 2011).

While reasoning about evolutionary change is diverse and complex, expert-novice reasoning patterns, and associated benchmarks of competency, may be simplified and modeled based on existing research (see Table 2). In this framework, novices are defined as those who tend to use exclusively naïve ideas, or both naïve ideas and key concepts of natural selection, in their explanations of evolutionary change (Authors, 2011). In contrast, 'emerging experts' are those whose explanations include key concepts of natural selection, but not naive ideas (Authors, 2011). Though emerging experts provide accurate conceptions of selective causation, they may fail to incorporate possible non-adaptive factors, which are more common in experts' models of evolutionary change (Authors, 2011). Experts may attribute change to *either* natural selection *or* genetic drift (e.g., Orr, 1998), or they may incorporate both selection *and* genetic drift (or other stochastic, non-deterministic processes) into their models of evolutionary causation (e.g., Ackermann and Cheverud, 2004). Considering that biologists employ either or both of these factors, it is reasonable to use these conceptual models as benchmarks for expert-like reasoning about evolution for life sciences students (Table 2, top row).

[Insert Table 2 about here]

Nonetheless, students have difficulty understanding and reasoning about evolution, and their naive ideas are well documented in the literature (e.g., Gregory, 2009). Though traditional perspectives of conceptual change have approached naive ideas as extinguishable or replaceable (Posner et al., 1982), others have documented the persistence of naïve ideas, while new, accurate scientific concepts are added to naive 'knowledge' frameworks (Vosniadou, Vamvakoussi & Skopeliti, 2008; Kelemen and Rosset, 2009). Indeed, students may assimilate scientific concepts learned in school into their pre-existing knowledge frameworks, unaware of any conflict between the two, thereby creating mixed or synthetic mental models of the phenomenon (Vosniadou et al., 2008; Authors, 2011). Based on prior studies, it is likely that most high school and undergraduate students are concentrated at the bottom end of our expert-novice continuum for evolutionary reasoning (where reasoning about evolutionary change either involves only naïve concepts and discarded historical ideas or comprises mixed models composed of both naive ideas and some accurate and/or discarded scientific concepts; see Table 2, bottom row).

Not all students neatly fit into these categories, however. Previous studies have shown that some undergraduate students are able to successfully reason about natural selection using accurate knowledge elements (key concepts) without employing any naive ideas (Nehm and Schonfeld, 2008; Nehm and Ha, 2011). Though prior work has not investigated non-adaptive reasoning specifically in relation to the aforementioned expert-novice continuum, the students from these studies would be considered 'emerging experts' rather than 'experts', because despite their lack of naïve ideas, they do not include stochastic, non-deterministic processes as possible mechanisms of evolutionary change in their explanations. Furthermore, the participants in these studies were enrolled in introductory biology courses for majors and, presumably, had limited exposure to stochastic mechanisms of evolutionary change. Whether these students needed more exposure to instruction about non-adaptive reasoning before integrating it into their mental frameworks of evolution remains to be determined. In short, our evolutionary competency framework is used to situate students' evolutionary reasoning sophistication.

Research Questions

In this study we explore if - and how - undergraduate students incorporate non-adaptive factors into their explanations of evolutionary causation. Specifically, we ask three questions: (1) Do students use non-adaptive factors to explain evolutionary change, and if so, does the frequency increase with increasing evolution coursework? (2) Is the use of non-adaptive reasoning patterns associated with greater knowledge of natural selection, or with fewer naive ideas? (3) How do students structure explanatory models of evolutionary change when both adaptive and non-adaptive factors are included?

Sample

We gathered data from undergraduate biology majors at a large, public, Midwestern research university in the United States. Fifty-five students from two cohorts were studied. The first cohort was from early in a college biology program (second semester introductory biology) and the second was from late in the program (an advanced organismal biology class with a prerequisite of an upper-division evolution class). The first cohort was exposed to basic evolution content (including non-adaptive factors such as genetic drift), and evolution was also considered by the course instructor to be a 'key theme'. The second cohort had extensive exposure to evolution (including selective and non-adaptive factors) in the introductory biology course, the advanced evolution course, and in the advanced organismal biology course. For brevity, we will refer to these two samples as 'majors' and 'advanced majors'. The majors sample consisted of 28 students (43% male, 57% female) with an average age 20.4 years. The advanced majors comprised 27 students (56% male and 44% female) with an average age of 21.9 years. The majority of students in both samples were White, non-Hispanic.

Methods

We used three methods to gather data on evolutionary reasoning patterns in the participants: (1) clinical oral interviews; (2) the open-response ACORNS assessment (Nehm, Beggrow, Opfer and Ha, 2012); and (3) the multiple-choice CINS test (Anderson, Fisher & Norman, 2002). Despite displaying psychometric problems (Battisti, Hanegan, Sudweeks &

Cates, 2010), the CINS is recognized as an instrument capable of generating valid inferences about general levels of students' evolutionary knowledge. Each item of the CINS has one correct response option for each question; therefore, the total score of the CINS instrument ranged from 0 to 20. While the original CINS paper suggests that it is a test only of natural selection knowledge, in fact it includes some questions about speciation, which is widely recognized as a macroevolutionary concept (Futuyma, 2008). For our sample of biology students, the reliability of CINS scores (measured with Cronbach's alpha) was 0.7.

The second instrument that we used was the newly developed open-response ACORNS (Assessing COntextual Reasoning about Natural Selection; Nehm et al., 2012). We used four isomorphic ACORNS items, standardized by familiarity: (1) How would biologists explain how a living mouse species without claws evolved from an ancestral mouse species that had claws? (2) How would biologists explain how a living lily species without petals evolved from an ancestral lily species that had petals? (3) How would biologists explain how a living snail species with teeth evolved from an ancestral snail species that lacked teeth? (4) How would biologists explain how a living grape species with tendrils evolved from an ancestral grape species that lacked tendrils?

The ACORNS is a test of both microevolutionary and macroevolutionary knowledge because it prompts students to explain the causes (e.g., natural selection) responsible for between-species (i.e., macroevolutionary) change. To score students' ACORNS responses, we utilized the published rubric of Nehm and colleagues (2010). This scoring rubric includes seven key conceptions and six naïve ideas. Key concept (KC) scores for each item ranged from 0 to 7, and naive idea scores for each item ranged from 0 to 6. The ACORNS responses were scored to consensus by two raters: a Ph.D. student in biology education and an evolutionary biologist.

ACORNS reliabilities (measured using Cronbach's alpha) were 0.8 for KCs, 0.6 for naive ideas, and 0.8 for non-adaptive reasoning (for detailed example scoring, see below).

The third approach that we used to explore students' evolutionary reasoning was clinical oral interviews. All 55 students were recruited as volunteers by the interviewer by e-mail as well as at the beginning and end of various class periods and laboratory sessions, and these participants reflected the performance distribution in the overall sample. All participants were offered USD $20 for their participation. Approximately 16 hours of interviews were audio recorded and transcribed. The majors sample comprised more than 10 hours of oral questioning (mean 19 minutes/student; range of 12-34 minutes). The advanced majors sample consisted of more than six hours of oral questioning (mean 15 minutes/student; range 8-24 minutes).

The interview protocol was comprised of two ACORNS items (identical to those on the written instrument) and two novel isomorphic items with taxa and traits of comparable familiarity (see Nehm et al., 2012). The two novel isomorphic items were included in the interview to minimize a potential testing effect (i.e., higher scores because of prior attempts to solve the same problem). While answering, students were prompted by the interviewer to elaborate on what they had said or to clarify what they meant by the words that they used. Follow-up questions included prompts such as 'Can you tell me more about X?' 'Can you explain what you mean when you use the word X?' and 'Can you tell me a little bit more about how X would happen, in general terms?' Interviews were analyzed by two raters and scored 0 for the absence of non-adaptive reasoning, KCs, or naïve ideas, and 1 for the presence of non-adaptive reasoning, KCs, or naïve ideas. An evolutionary biologist and a biology education Ph.D. student evaluated all oral interviews. Initial inter-rater reliabilities were 0.75 for oral interview scoring, and all discrepancies were subsequently resolved by deliberation. Consensus scores were used in all subsequent analyses.

We calculated descriptive statistics for the two samples (majors and advanced majors) and compared participant performance for all measured variables between the two groups using t tests. This information is useful for aligning our student sample with previously studied student samples in evolution education that used the same instruments (e.g., CINS). We also calculated Pearson correlation coefficients to examine putative interrelationships among measured variables from the multiple-choice CINS test, the open-response ACORNS test, and the clinical interviews. Variables included (1) the number of correct CINS scores; (2) the number of written key concepts of natural selection documented in the ACORNS; (3) the number of written naïve ideas identified in the ACORNS; (4) the number of written non-adaptive evolutionary factors in the ACORNS; (5) the number of mentions (or 'naming') of non-adaptive evolutionary factors in the clinical interviews; (6) the number of key concepts of natural selection in the interviews; and (7) the naïve ideas mentioned in the interviews. Pearson correlation coefficients were calculated in PASW v. 18.

Finally, we used qualitative methods to examine the structure of students' reasoning patterns. We examined transcripts of the interviews for (1) patterns of how students incorporated stochastic causal factors, such as genetic drift, into their explanations and (2) whether they conceptualized non-adaptive factors as alternatives to selection or as synergistic 'forces' of change. The purpose of including oral interviews was not only to validate inferences derived from students' written answers, but also to create a holistic snapshot of students' evolutionary reasoning patterns.

Results

Associations among evolutionary reasoning elements. Pearson correlation analyses indicated that non-adaptive reasoning (NAR) interview scores were significantly correlated with NAR scores from the open-response ACORNS, but not with any scores from the multiple-choice CINS (Table 3). Specifically, both 'mentioning' and 'scientifically explaining' NAR scores from the interviews showed strong and significant associations with ACORNS NAR scores ($r = 0.86$, $p < 0.01$ and $r = 0.84$, $p < 0.01$, respectively). Higher NAR 'explaining' scores were *not* significantly associated with greater key concept (KC) scores for the ACORNS ($r = 0.04$, n.s.) or with higher CINS scores ($r = 0.08$, n.s.). Thus, NAR appears to be a somewhat distinct reasoning pattern from selective reasoning.

[Insert Table 3 about here]

Explanatory elements used by majors and advanced majors. We compared the types and frequencies of conceptual elements used in students' explanations of evolutionary change (Figure 1). We found slight increases in students' accurate knowledge elements (KCs, CINS scores) between the samples of majors and advanced majors (using both interview data and ACORNS data). However, these differences were not significant (CINS: $t_{47}= 1.33$, p. > 0.05; ACORNS KC: $t_{53}= 0.63$, p. > 0.05; Interviews KC: $t_{53}= 1.31$, p. > 0.05). The types and frequencies of naïve ideas also did not differ appreciably between the two cohorts (ACORNS: $t_{53}= 0.88$, p. > 0.05; Interviews: $t_{53}= 0.24$, p. > 0.05). In contrast, NAR was much more frequent in the sample of advanced majors, as measured by both the clinical interviews and the open-response ACORNS test (the CINS does not include NAR options) (Figure 1). However, NAR patterns were only significant for those students who 'mentioned' NAR in their interview responses ($t_{28}= 2.25$, p. < 0.05, $d = 0.61$; Cohen, 1988).

[Insert Figure 1 about here]

Explanatory elements used in the ACORNS and clinical interviews. The KCs of *variability, differential survival, and limited resources* were the three most frequently used in the ACORNS responses and in the clinical interviews (Figure 2). However, the relative frequencies of these concepts differed for each assessment. In the interviews, *variability* was the most frequent, followed by *limited resources* and then *differential survival*. Overall, KCs were used more often in the interviews than in the written ACORNS instrument. In the written ACORNS assessment, *differential survival* was used most frequently, followed by *variability* and then *limited resources*. Relative proportions of the other four KCs were similar between written and oral responses, and in order of relative frequencies were *hereditability*, *change in the frequency of a variant in a population*, c*ompetition* and *hyper-fecundity. Hyper-fecundity* was the least frequently applied KC; it was never used in the ACORNS and was only used once in the interviews. Overall, KCs were greater in number in the interviews than in the ACORNS, which is expected given the greater time allotted to oral questioning. Students spent almost twice as much time answering interview questions as they did answering ACORNS items (ACORNS: M=10.9, SD = 7.2 minutes; Interview: M=21.9, SD = 7.7 minutes (Figure 2). Overall, however, there was good correspondence between measures derived from the two methods.

[Enter Figure 2 about here]

Similar to the KC patterns that we documented, naive ideas were more abundant in interview responses than in the ACORNS responses (Figure 2). In both oral interviews and in written responses, however, naïve ideas were much more variable across items than KCs. The most frequent naïve idea used in both the interviews and in the ACORNS was *needs/goals*. The naive ideas *use/disuse, energy,* and *pressure* were about equally common in ACORNS responses. *Adapt* was the least common naive idea found in the ACORNS responses. In the interviews, the second most common naive idea was *pressure*, while *adapt* and *energy* were equally the next most frequent. *Use/disuse* was the least common naive idea used during interviews. One notable pattern relating to NAR in both the ACORNS and interviews, was that it was used in the first item more often than in other items, and it was inconsistently applied across items (Figure 2, top row). In short, the interviews tended to elicit greater frequencies of ideas, but not different ideas, than those revealed in the ACORNS. NAR was rarely used regardless of the method of detection used, and naïve ideas were less common than KCs.

Holistic reasoning patterns. Interview data revealed that all of the students who incorporated non-adaptive reasoning (NAR) into their explanations of evolutionary change (eight out of 55) presented NAR in the form of 'genetic drift', 'bottleneck', or 'founder's effect' (See Table 4). Moreover, in all cases students represented NAR as an alternative to natural selection. This finding is in line with how many evolutionary biologists conceptualize the two concepts, where evolutionary change is attributed to either natural selection *or* genetic drift (e.g., Orr, 1998).

[Insert Table 4 about here]

While at first glance the observation that students discussed genetic drift as an alternative to natural selection was indicative of expert-like thinking (see Introduction), follow-up questions during the interviews often revealed that many students poorly understood the concepts of genetic drift and/or natural selection. Indeed, only one student (Participant G from the advanced majors sample) accurately and consistently used both genetic drift and natural selection as possible mechanisms for evolutionary change across the four interview items (see below).

Participant G: …another way would be more like genetic drift oriented, where, the presence or absence of a tail doesn't matter very much. So, maybe it has no fitness ramifications in survivability and it just kinda [sic] fluctuates based on just randomness, you know, statistics, until eventually you've got all the opossums don't have tails. That sort of random stuff does happen, sometimes, and it brings genes towards fixture that way. But, I would say, cause it, I would say most biologists would go with either natural selection or gene drift to explain the loss of tail over time in opossums.

Interviewer: So when you're talking about the gene drift option, um, in that scenario, there would be no selection or there would be selection or…

Participant G: There'd be no selection in like a more random gene drift. That would be, you know, in whatever scenario these opossums are living in the presence or absence of a tail has absolutely no impact on survivability, but the variation still exists. So, you know, due to, you know, maybe one opossum got lucky and found a lot of females in a small population or, you know, it just happened. Or maybe there was like, a tornado, that came and killed a bunch of opossums that had tails or any other number or random, uncontrollable events that don't really have anything to do with the tail or lack of tail giving some sort of selective advantage. I mean, that's how most, like, you're flipping a coin just to see which one comes out on top. Or lots and lots of coins. Thousands and thousands of times.

Interviewer: So would you describe, um, the alternative, so we have the two options, the gene drift or the selection option, and gene drift you say is completely random, would selection be random or no?

Participant G: Selection would not be random.

Interviewer: Ok.

Participant G: Selection would be you've got some sort of reproductive advantage given by the lack of tail and that's to the point where it's, I mean, you could predict which opossums would do better and which opossums would do worse due to the presence or absence of the tail. So that's like, there's a, you know, like one thing I could think of would be opossums have a naked tail and in colder climates that could be a great way to lose heat, so maybe, you know, these opossums are living in a cold part of the world and all the opossums that have this large exposed naked tail lose a bunch of energy because they're not insulated and the opossums that don't have a tail save that energy and put it into reproduction or offspring or whatever and they end up coming out with better results. So that's less ran-, that's not random, that's, you know, it's predictable, and there's, like, the natural selection, you know, force, that's kinda [sic] acting on it. Where there's like a quantifiable difference in the fitness, and like the vitality and viability of the gene.

A second student from the advanced majors sample presented 'genetic drift' as one possible solution for the fourth interview item about cactus spines (see Methods).

Participant E: I guess that could be similar with either, like a genetic mutation or maybe a genetic drift and, uh, just, could also have to do with, uh, like being an anti-predatory defense to, uh, protect it since it's in a harsh environment already, they kinda [sic] have to guard themselves from anything that's going to get water out of them.

Interviewer: You said that it could be a genetic mutation or genetic drift, can you explain what you mean by genetic drift?

Participant E: I always forget the definition of genetic drift. Um, it's kinda [sic] just like, uh, a swing towards one extreme instead of where it was before, but I guess, so I guess that kind of takes away from what I was going towards, I guess, kinda [sic] contradicted myself.

Interviewer: What's swinging? What is that is swinging towards one extreme?

Participant E: Uh, just kinda [sic] like, the genetic makeup or the, uh, actual structures that are present are kinda [sic] more of a shift from one to another based on the pressures that they're getting from the environment or from, from other species around them.

Interviewer: When you say structures, do you mean, um, what do you mean by that?

Participant E: Um, sorta [sic] like the actual spines, in this case, or uh, the the [sic] cells within, that are used to maintain the water and the moisture and hold it.

Interviewer: And how, how does that process of swinging towards one of the extremes or another come about?

Participant E: Um, I guess that would be more, type of a, kind of, more of natural selection, where they're, where if they don't change then they're just going to die out, so without, if the, if they didn't develop the spines, there'd be no way for them to completely protect themselves, they would, without spines they'd probably evolved in another way and came up with some sort of poison or something else that would have deterred predators and stuff from eating them.

While participant E mentions the term 'genetic drift', his explanation for this term was scientifically inaccurate (the student confuses genetic drift with directional selection, and he was therefore given a scientific NAR score of '0'). His answer demonstrates that a student's use of a scientific term is not always indicative of scientific understanding of a concept (cf. Rector, Nehm and Ha, 2011). This situation also occurred with students' use of the terms 'bottleneck' and 'founder effect'. When prompted to explain the meanings of these terms the students were unable to provide an accurate or clear definition (see below).

Interviewer: …how would biologists explain how a living opossum species without a tail evolved from an ancestral opossum species that had a tail?

Participant A: Um, hold on, well it could also, it could be the bottleneck effect or the founder's effect, like, the ancestor species could have had like a nat-, disaster, like a, calamity, so only a few of the, uh, individuals survived and those individuals like maybe had like a dwarf gene, or like a, like a gimp tail I guess. And then eventually it started being more prevalent in the, in the new species that started from there. Or, just, um, reproductive success…

Interviewer: …And you talked about how, uh, back with this bottleneck effect or this founder effect…

Participant A: Mm hm.

Interviewer: That there were individual, a few individuals that had a, a dwarf tail or, um, gimp tail…

Participant A: …Yeah.

Interviewer: Um, how did that, how did they have that tail, or how did they get that different tail?

Participant A: Um, maybe like survivors from the calamity had like some, uh, maybe they did, maybe they didn't, like they could have expressed that gene in like, uh, um, uh, in like a repressed way, than having like a full on gene or maybe they had the full gene and then, like, eventually, wait can you repeat your question? Sorry.

(General confusion ensues and then Participant A provides an adaptive explanation)

Overall, clinical interviews corroborated our statistical findings that knowledge of NAR is not associated with greater understanding of natural selection, or with fewer evolutionary naive ideas. Similarly, in clinical interviews and in ACORNS responses, most students inconsistently applied NAR, or lacked what has been termed 'knowledge coherence' (Kampourakis & Zogza, 2009).

Evolutionary Reasoning Competencies in the two cohorts. We found that on the spectrum of novice to expert reasoning about evolutionary change, the majority of our students (both majors and advanced majors) fell into the *novice* category (Table 2, Figure 3). Recall that the novice category was characterized by naïve or naïve + scientific reasoning. Based upon the written explanations, only one student from the majors sample held a purely *naïve* model, while no students from the advanced majors cohort held such models. In the interviews, none of the students from either cohort were found to exhibit purely naïve models. Most students exhibited mixed models comprised of both naïve ideas and KCs (Figure 3). Written responses indicated that 16 majors and 14 advanced majors held mixed models, whereas interview responses indicated that both cohorts had 19 students exhibiting mixed models. Interestingly, a small number of students from each cohort displayed mixed models while also incorporating non-adaptive reasoning into their explanations (ACORNS: two from the majors sample, two from the advanced majors sample; Interviews: one from the majors sample, four from the advanced majors) (Figure 3). There were a number of students categorized as *emerging experts* based on their written explanations (eight majors, ten advanced majors). During the interviews, in contrast, seven majors exhibited purely adaptive models, although only two advanced majors used KCs exclusively (Figure 3). Nevertheless, a small number of advanced majors fell into the *expert* category (that is, students explaining evolutionary change using both adaptive and stochastic conceptual models; ACORNS: two students; Interviews: three students). No students from the majors sample reached the *expert* level.

[Enter Figure 3 about here]

It was challenging to unambiguously situate a few students along our novice-expert continuum. For instance, one student from the advanced majors sample displayed a complex mixed model (naive ideas + KC + NAR) in the interview (Participant F), but not in the written assessment. Based on her ACORNS responses, we placed Participant F in the adaptive vs. stochastic model of reasoning even though she had provided explanations using NAR exclusively. She was placed in this category because she had used key concepts (*variability* and *change of population*) in her explanations of genetic drift. However, her response demonstrated that some key concepts were not specific to natural selection; in fact, KCs such as *variability* and *change of population* are necessary for explaining evolutionary change by stochastic processes as well as deterministic processes. Accordingly, this student was labeled as having an expert-like model of evolutionary reasoning. However, placing students along a continuum of evolutionary reasoning competency was straightforward in most cases. Classifying our two student samples revealed that, regardless of the assessment method or amount of biology coursework completed, most students failed to reach expert-like levels of reasoning about evolutionary change.

Discussion

While science education research has produced a large body of work investigating the learning, teaching and assessment of natural selection (e.g., Bishop and Anderson, 1990; Demastes et al., 2001; Settlage, 1994; Nehm and Ha 2011; Nehm and Schonfeld, 2008), strikingly few studies have focused on students' thinking about non-adaptive evolutionary factors such as genetic drift, despite its important role in experts' empirical tests and theoretical models of evolutionary change (e.g., Ackermann and Cheverud, 2004; Lande, 1976; Orr, 1998; Parker and Maynard Smith, 1990). This gap in the literature motivated our exploration of non-adaptive reasoning patterns (NAR) in undergraduate students, and whether such patterns differed between cohorts of introductory and advanced biology students exposed to varying degrees of non-adaptive content. We used three different methodologies to explore students' NAR: the multiple-choice CINS, which ignores NAR but measures understanding of natural selection and speciation; the ACORNS, a constructed-response format test that captures students' evolutionary explanations across contextual features; and extended clinical oral interviews. We used these three different methods to rigorously and holistically determine how, and to what degree, students used stochastic and non-deterministic factors to explain evolutionary change, and how their explanatory models were constructed.

Students' use of NAR. Given Gould and Lewontin's (1979) widely-cited criticisms about biologists' over-reliance on exclusively adaptive factors to explain evolutionary change, and contemporary biologists' use of hypothesis tests that explore the relative contributions of non-adaptive mechanisms such as genetic drift in evolutionary change, the goal for biology students should be to understand and apply the mechanisms that the field of evolutionary biology currently uses to account for organismal diversity through time and space (Gould, 2002). In short, students' competency in evolutionary reasoning should be measured by their ability to consider both drift and selection as possible causal mechanisms in their explanations for trait change. What our study reveals is that the vast majority of students in our sample have not reached this competency benchmark.

Student use of non-adaptive reasoning is not associated with greater knowledge of natural selection or with fewer naive ideas about evolution. Most students in the majors and advanced major samples used mixed models of evolutionary reasoning, implying that despite increased instruction in evolution, most students failed to reach expert-like conceptualizations of evolutionary change. Instead, their responses suggest that they simply added new concepts (e.g. *genetic drift* or *differential survival*) into their existing naïve explanatory frameworks and regardless of the amount of biology coursework completed, expressed both KCs and naive ideas at comparatively the same frequency when cued to reason about evolutionary change (Figure 3). Nevertheless, advanced majors' relatively greater use of NAR suggests that increased exposure to instruction about stochastic processes, such as genetic drift, is in fact associated with increases in students' use of NAR in their evolutionary explanations, (albeit not at desired magnitudes; see Figure 1).

The structure of students' explanatory models of evolutionary change. When students did use NAR in their explanatory models of evolution, they either used it within a mixed model of reasoning, or they used it as a competing mechanism of evolutionary change (i.e., in an expert-like model of adaptive vs. stochastic reasoning; Figure 3). Additionally, most students did not consistently apply NAR across items (Figure 2), suggesting that the association of this concept within students' evolutionary reasoning framework is not theory-like (Vosniadou et al., 2008) or was selectively cued across items differing in surface features (Nehm et al., 2012; Nehm and Ha, 2011). Indeed, only two of the three students from the advanced majors cohort who did use NAR, did so consistently and expressed an adaptive vs. stochastic model of reasoning about evolutionary change. This corroborates previous work suggesting that expert-like evolutionary reasoning models include more stable associations of concepts within conceptual reasoning frameworks (Nehm and Ha, 2011).

Those students who reached an expert-like model of evolutionary causation adopted models of stochastic vs. adaptive change, but none of them expressed integrative causal models (i.e., the top level of Table 2). For example, Participant G (see excerpt above) was a student from the advanced majors cohort who included both natural selection and NAR across all four items in both the ACORNS written assessment and in the interview. Participant G employed what we term an expert-like model of reasoning, although one that frames adaptive causation as *competing* with stochastic causation of evolutionary change. Another student, Participant F, provided only NAR in the ACORNS explanations, yet also discussed NAR and selection as possible alternative causal mechanisms of evolutionary change in the interviews. It is interesting that the written format of the ACORNS elicited NAR only, while the interviews elicited both possible mechanisms. However, this seems to be an anomaly and not significant enough to warrant any adjustments to our framework of novice-expert reasoning about evolutionary change (Figure 2). Regardless of assessment format, this student reached expert levels of reasoning.

Implications for teaching and learning. Our study demonstrates that current approaches to teaching genetic drift and other non-deterministic processes may not be effective at helping the majority of students attain expert-like models of evolutionary causation. Students' models of evolutionary change in our sample were overwhelmingly composed of a combination of naive ideas and key concepts of natural selection (Figure 2). Not only is this the case for majors in introductory courses; it is also true of advanced majors who have successfully completed an entire course in evolutionary biology. Clearly, exposure to genetic drift does not appear to be sufficient for inducing students to accommodate NAR into their mental models of evolutionary change and build expert-like explanatory models (i.e., Table 2, top row). This finding suggests that teachers should move beyond definitions or simulations of genetic drift (e.g., Staub, 2002) and illustrate to students how genetic drift is conceptually structured within models of evolutionary causation. Teachers could present cases of how evolutionary biologists currently use both genetic drift and natural selection to test alternative hypotheses and build explanatory models of evolutionary change. Such examples may facilitate more advanced perspectives on stochastic and adaptive causal factors in evolutionary biology (e.g. Orr, 1998; Ackermann and Cheverud, 2004).

Our findings also suggest that NAR and scientifically accurate selective reasoning are distinct patterns of thinking about evolution. This raises the question of whether the lack of an association among NAR, KCs, and naïve ideas is a product of the way these evolutionary concepts are taught, or if it is an intrinsic way of thinking about evolution. For instance, perhaps students fail to use NAR in their evolutionary explanations because they view the lack of emphasis on drift in the classroom as indicative of its relatively low level of importance in evolutionary change. On the other hand, students may fail to use NAR in their evolutionary explanations because the concept is more difficult to accommodate into existing cognitive frameworks compared to selective reasoning concepts. Regardless, based on students' response patterns, it appears that they have not sufficiently accommodated NAR into their evolutionary reasoning frameworks, even after advanced instruction in evolution. Rather, interviews and open-response assessments clearly indicate that most students at introductory and advanced levels of instruction assimilate NAR concepts with KCs and naive ideas of selective reasoning into unstable (i.e., non-coherent), naïve models of evolutionary causation (cf. Kampourakis & Zogza, 2009; Nehm and Ha, 2011). It is important to note, however, that student response patterns may not always mirror cognitive processes, and our interpretations are constrained by the methods we used to uncover student thinking and the sample that we studied.

The finding that *no* students used selective + stochastic causation models in their explanations of evolutionary change raises the question of whether this, too, is a product of teaching experiences or a default approach to thinking about evolution. Stochastic factors such as genetic drift often receive minor instructional focus and are presented as an alternative model to natural selection (e.g., Linhart, 1997). It is possible that standard approaches prevent students from conceptualizing cases in which stochastic processes work in tandem with selective processes to generate patterns of genotypic and phenotypic change. Such integrated causal models (selective + stochastic) are perhaps more complex, and thereby may be beyond the grasp of students still struggling to restructure common naive ideas. Indeed, it may require more comprehensive evolution instruction using extensive theoretical and experimental examples. However, it is important to note that some evolution experts continue to use an adaptive vs. stochastic model of evolutionary change in their research programs (e.g. Orr, 1998), and so it is possible that a selective + stochastic model is simply a less common, alternative framework for experts within evolutionary biology. Further work is needed to explore this issue.

Overall, our study highlights the fact that NAR receives very little attention in the science education or evolution education literature. Future research exploring novice and expert understanding and application of NAR would help to reveal what a more expert-like model of evolutionary reasoning would look like and could inform the design of evolution learning progressions (cf. Catley et al. 2005).

Implications for assessment. Assessments are central to helping teachers foster meaningful science learning (NRC, 2001b). However, it is imperative that those assessments meet quality control criteria established by the educational measurement community (AERA, 2004). Measurement instruments, among other things, must comprehensively assess all facets of a well-defined construct (Neumann, Neumann, and Nehm, 2011). In the domain of evolutionary biology, natural selection and genetic drift are the two most important causal factors that biologists use to explain evolutionary change (Orr, 1998). Thus, to measure evolutionary thinking, instruments must provide the opportunity for students to explain change using natural selection, genetic drift, or combinations of drift + selection. Currently, there are no such instruments.

One important consideration for helping teachers to understand students' thinking about evolution is to employ instruments that are capable of capturing progress in students' conceptual growth. As we argue, measuring progress in evolutionary reasoning requires the consideration of both stochastic and deterministic factors in evolutionary causation, in addition to common naive ideas about evolutionary change (Table 2). Many widely used instruments, nevertheless, ignore the possibility that non-adaptive factors (such as genetic drift) could contribute to patterns of evolutionary change that the instrument scenarios present (e.g., Bird beak evolution in Anderson et al.'s CINS instrument). Thus, teachers across the educational hierarchy must develop and deploy instruments that include the measurement of NAR in formative and summative assessments of evolutionary thinking.

Multiple-choice assessments are often a popular choice among instructors considering the time and expertise needed to develop and grade open-ended items. However, multiple-choice items rarely provide mixed model options (that is, not just right or wrong options), despite the fact that such reasoning models may be very common in samples (as we found in our study). If we want our assessments to fulfill their purposes, we must consider the large proportion of students who have mixed models for explaining evolutionary change and revise our multiple choice assessments to reflect this well established finding. No multiple-choice instruments to our knowledge allow for the measurement of mixed models.

Study limitations and implications for future work on non-adaptive reasoning. Although we were able to describe *how* students reason about evolutionary change and their use of NAR, we were not able to determine *why* students did not incorporate NAR into their models of evolutionary causation. Because all students in our sample were assigned readings that covered genetic drift, and had the opportunity to listen to lectures that included the topic of genetic drift, the lack of NAR in student explanations could be due to poor teaching of the topic. Indeed, although the curriculum included material on genetic drift, we do not know *how* drift was taught to students. Additionally, it is unclear whether teachers incorporated developmental constraints, or other lesser-known non-adaptive factors (see Gould, 2002), into their teaching, or whether the curricula included them at all. It is possible that other student cohorts exposed to instruction on drift in a more integrated and sophisticated manner would show different patterns than those observed in our sample. Thus, our findings may not generalize to other student samples. Future studies should explore the relationship between instruction and NAR to determine whether specifically addressing NAR (beyond genetic drift) through instruction, or how teaching NAR in different ways, influences students' use of NAR in their evolutionary explanations.

We have described how evolutionary biologists employ (1) selective factors, (2) stochastic factors, or (3) a combination of both factors to explain evolutionary causation (Futuyma, 2009). We consider biologists' explanations as reasonable expert-level targets for students in life sciences courses. However, it is important to note that philosophers of science may not agree with how evolutionary biologists' perceive and study selective and stochastic mechanisms of evolution. Though evolutionary biologists appear to consider natural selection and genetic drift to be distinct concepts and attribute evolutionary change to either one (e.g. Orr, 1998) or both (e.g., Ackermann and Cheverud, 2004), philosophers of science are not in agreement as to whether in fact drift and selection are distinct concepts or not (e.g., Beatty, 1984; Sober, 1984; Matthen and Ariew, 2002; Millstein, 2002). This issue clearly requires further attention. Interviews with evolutionary biologists that explore their perceptions of selective and non-adaptive causal mechanisms could provide insights into experts' perspectives regarding evolutionary causation. Additionally, both evolutionary biologists and philosophers of biology would benefit from an open dialogue about the causal factors of evolutionary change; such a conversation would be fruitful, not just for the progression of their respective fields, but also for the advancement of science education research and practice.

References

American Association for the Advancement of Science. (2011). American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education, Washington, DC: http://visionandchange.org

Ackermann, R.R. & Cheverud, J.M. (2004). Detecting genetic drift versus selection in human evolution. *Proceedings of the National Academy of Sciences, 101*, 17946-17951.

AERA. (2004). American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME). (2004). Standards for educational and psychological testing. Washington, DC: AERA.

Anderson, D.L., Fisher, K.M., & Norman, G.J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching, 39,* 952-978.

Battisti, B.T., Hanegan, N., Sudweeks, R. & Cates, R. (2010). Using item response theory to conduct a distracter analysis on Conceptual Inventory of Natural Selection. *International Journal of Science and Mathematics Education, 8*, 845-868.

Beatty, J. (1984). Chance and natural selection. *Philosophy of Science, 51*(2), 183-211.

Bishop, B.A. & Anderson, C.W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of Research in Science Teaching, 27*(5), 415-427.

Campbell, N.A., Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky P.V., & R.B. Jackson. (2008). *Biology*. 8th ed. New York: Pearson Benjamin Cummings.

Carpenter, P., Bullock, A., & Potter, J. (2006). Textbooks in teaching and learning: the views of students and their teachers. *Brooks eJournal of Learning and Teaching*. *2*, 1–10.

Catley, K., Lehrer, R. & Reiser, B. (2005). Tracing a prospective learning progression for developing understanding of evolution. Paper commissioned by the National Academies Committee on Test Design for K-12 Science Achievement. Washington, DC: National Academies. http://www7.nationalacademies.org/bota/Test_Design_K-12_Science.html.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Conway Morris, S. 2003. *Life's Solution: Inevitable humans in a Lonely Universe*. Cambridge: Cambridge University Press.

Cuseo, J. (2007). The Empirical Case Against Large Class Size: Adverse Effects on the Teaching, Learning, and Retention of First-Year Students. *The Journal of Faculty Development, 21*(1), 5-21.

Darwin, C. (1859). *On the Origin of Species By Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray.

Dagher, Z.R. & BouJaoude, S. (1997). Scientific views and religious beliefs of college students: The case of biological evolution. *Journal of Research in Science Teaching, 34*(5), 429-445.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Demastes, S.S., Good, R.G., and Peebles, P. (1995). Students' conceptual ecologies and the process of conceptual change in evolution. *Science Education, 79*(6), 637-666.

Demastes, S.S., Good, R.G., and Peebles, P. (1996). Patterns of conceptual change in evolution. *Journal of Research in Science Teaching, 33*(4), 407-431.
Demastes, S.S., Settlage, J. and Good, R.G. (1995). Students' conceptions of natural selection and its role in evolution: cases of replication and comparison. *Journal of Research in Science Teaching, 32*(5), 535-550.
Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher, 35*(3), 125 – 129.
Ferrari, M. and Chi, M.T.H. (1998). The nature of naïve explanations of natural selection. *International Journal of Science Education, 20*(10), 1231-1256.
Fisher, R.A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh, 42*, 321-341.
Fosnot, C. T. (1996). Constructivism: A psychological theory of learning. In *Constructivism: Theory, Perspectives and Practice*, ed. C. T. Fosnot, (pp. 8–33). New York: Teachers College Press.
Freeman, S. (2005) *Biological Science*. 2nd ed. Upper Saddle River: Pearson Prentice Hall.
Futuyma, D.J. (1998). *Evolutionary Biology*. 3rd ed. Sunderland: Sinauer Associates, Inc.
Futuyma, D.J. (2009). *Evolution*. 2nd ed. Sunderland: Sinauer Associates, Inc.
Gould, S.J. (2002). *The Structure of Evolutionary Theory*. Cambridge: Belknap Press.
Gould, S.J. and Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London, B, 205,* 581-598.
Gregory, T. R. (2009). Understanding Natural Selection: Essential Concepts and Common Misconceptions*.* *Evolution Education and Outreach, 2*(2), 156-175.
Hagedoorn, A.L. & Hagedoorn, A.C. (1921). *The Relative Value of the Processes Causing Evolution.* The Hague: Martinus Nijhoff.
Keil FC, &Wilson RA. (2000). *Explanation and Cognition*. Cambridge, MA: MIT Press.
Kelemen, D. & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111*(1), 138-143.
Kampourakis K, & Zogza V. (2009). Preliminary evolutionary explanations: a basic framework for conceptual change and explanatory coherence in evolution. *Science & Education, 18*(10):1313–40.
Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature, 217*, 624-626.
Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution, 30*(2), 314-334.
Linhart, Y.B. (1997). The teaching of evolution: We need to do better. *BioScience, 47*(6), 385-391.
Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. *Nature Reviews 8*, 803-813.
Maret, T.J. & Rissing, S.W. (1998). Exploring genetic drift & natural selection through a simulation activity. *The American Biology Teacher*, *60*(9), 681-683.
Matthen, M. & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *The Journal of Philosophy, 99*(2), 55-83.
Mayr, E. (1983). How to carry out the adaptationist program? *The American Naturalist, 121,* 324-334.
Millstein, R.L. (2002). Are random drift and natural selection conceptually distinct? *Biology and Philosophy, 17*, 33-53.

National Association of Biology Teachers. (2011). NABT Position Statement on Teaching Evolution. Accessed February 18, 2012, http://www.nabt.org/websites/institution/index.php?p=92.

National Research Council. (1996). *National science education standards.* Washington, D.C: National Academy Press.

National Research Council (2001a). *Investigating the Influence of Standards: A Framework for Research in Mathematics, Science, and Technology Education*. Washington, D.C.: National Academy Press.

National Research Council (2001b). *Knowing What Students Know*. Washington, D.C.: National Academy Press.

Nehm, R.H., Beggrow, E., Opfer, J., and Ha, M. (2012). Reasoning About Natural Selection: Diagnosing Contextual Competency Using the ACORNS Instrument. *The American Biology Teacher, 74*(2), 92-98.

Nehm R.H., Ha M, Rector M, Opfer J, Perrin L, Ridgway J, & Mollohan, K. (2010). Scoring guide for the open response instrument (ORI) and evolutionary gain and loss test (EGALT). Technical Report of National Science Foundation REESE Project 0909999. Accessed February 18, 2012, http://evolutionassessment.org/Evo_Assessment/Scoring.html.

Nehm, R.H., and Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching.* 48(3):237–256.

Nehm, R.H., Kim, S.Y., & Sheppard, K. (2009). Academic preparation in biology and advocacy for teaching evolution: Biology versus non-biology teachers. *Science Education, 93*(6), 1122–1146.

Nehm, R.H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*(3), 263–272.

Nehm, R.H. & Schonfeld, I.S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education, 18*(5), 699 – 723.

Nehm, R.H. and Schonfeld, I. (2008). Measuring knowledge of natural selection: a comparison of the CINS, and open-response instrument, and oral interview. *Journal of Research in Science Teaching, 45*(10): 1131-1160.

Neumann, I., Neumann, K., Nehm, R.H. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a Nature of Science Test. *International Journal of Science Education, 33*(10), 1373-1405.

Nickels, M.K., Nelson, C.E. & Beard, J. (1996). Better biology teaching by emphasizing evolution & the nature of science. *The American Biology Teacher, 58*(6), 332-336.

Orr, H.A. (1998). Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, 149, 2099-2104.

Parker, G.A. and Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature 348,* 27-33.

Piaget, J. (1964). Cognitive development in children: Piaget, development and learning. *Journal of Research in Science Teaching*, *2*(3), 176-186.

Pigliucci, M. and Müller, G.B. (2010). Elements of an extended evolutionary synthesis. In M. Pigliucci and G.B. Müller, (Eds.), *Evolution, the Extended Synthesis* (pp. 3-17). Cambridge, MA:MIT Press.

Posner, G.J., Strike, K.A., Hewson, P.W., and Gertzog, W.A. (1982). Accommodation of a

scientific conception: Towards a theory of conceptual change. *Science Education*, *67*(4), 489-508.
Provine, W.B. (1971). *The Origins of Theoretical Population Genetics*. Chicago: The University of Chicago Press.
Randall, L. (2011). Knocking on Heaven's Door: How Physics and Scientific Thinking Illuminate the Universe and the Modern World. Ecco Publishers.
Rector, M. A., Nehm, R. H., and Ha, M. (2011). Lexical ambiguity in evolutionary discourse: What students say versus what students mean in explanations of evolutionary change. Paper presented at National Association for Research in Science Teaching.
Savage, J.M. (1969). *Evolution*. 2nd ed. New York: Holt, Rinehart and Winston, Inc.
Settlage, J., Jr. (1994). Conceptions of natural selection: A snapshot of the sense-making process. *Journal of Research in Science Teaching*, 31(5), 449-457.
Sinatra, G.M., Brem, S.K., and Evans, E.M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education & Outreach, 1*, 189-195.
Sober, (1984). *The Nature of Selection: Evolutionary Theory in Philosophical Focus.* Cambridge, MA: MIT Press.
Staub, N.L. (2002). Teaching evolutionary mechanisms: Genetic drift and M&M's®. *BioScience*, *52*(4), 373-377.
Vosniadou, S. (1999). Conceptual change research: State of the art and future directions. In W. Schnotz, Vosniadou, S. , Carretero, M. (Ed.), *New Perspectives on Conceptual Change* (pp. 3-14). Pergamon: Oxford.
Vosniadou, S., Vamvakoussi, X.,& Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp.3–34). New York: Routledge.
Wallace, A.R. (1871). *Contributions to the Theory of Natural Selection*. 2nd ed. New York: Macmillan and Co.
Wandersee, J.H., Mintzes, J.J. and Novak, J.D. (1994). Research on alternative conceptions in science. In D.L. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning*, pp. 177-210. New York: Macmillan Publishing Company.

Tables and Figures

Table 1. Definitions of natural selection and genetic drift from glossaries of introductory college biology textbooks.

Word	Definition	Textbook
Natural Selection	*A process in which organisms with certain inherited characteristics are more likely to survive and reproduce than are organisms with other characteristics.*	Campbell, N.A., Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky P.V., and R.B. Jackson. (2008). Biology. 8th ed. New York: Pearson Benjamin Cummings.
Natural Selection	*The process by which individuals with certain heritable traits tend to produce more surviving offspring than do individuals without those traits, resulting in a change in the genetic makeup of the population. A major mechanism of evolution.*	Freeman, S. (2005) Biological Science. 2nd ed. Upper Saddle River: Pearson Prentice Hall.
Natural Selection	*The differential survival and/or reproduction of classes of entities that differ in one or more characteristics; the difference in survival and/or reproduction is not due to chance, and it must have the potential consequence of altering the proportions of the different entities to constitute*	Futuyma, D.J. (1998). Evolutionary Biology. 3rd ed. Sunderland: Sinauer Associates, Inc.

	natural selection. Thus natural selection is also definable as a partly or wholly deterministic difference in the contribution of different classes of entities to subsequent generations. Usually the differences are inherited. The entities may be alleles, genotypes or subsets of genotypes, populations, or in the broadest sense, species.	
Genetic Drift	*A process in which chance events cause unpredictable fluctuations in allele frequencies from one generation to the next. Effects of genetic drift are most pronounced small populations.*	Campbell et al. (2008). Biology. 8th ed. New York: Pearson Benjamin Cummings.
Genetic Drift	*Any change in allele frequencies due to random events. Causes allele frequencies to drift up and down randomly over time, and eventually can lead to the fixation or loss of alleles.* *Non-adaptive change (e.g., genetic drift) occurs when the frequency of a trait increases or decreases because of stochastic factors, regardless of whether the trait confers an advantage, disadvantage, or is neutral with respect to survival or reproduction (Freeman,*	Freeman, S. (2005) Biological Science. 2nd ed. Upper Saddle River: Pearson Prentice Hall.

	2005).	
Genetic Drift	*Random changes in the frequencies of two or more alleles or genotypes within a population*	Futuyma, D.J. (1998). Evolutionary Biology. 3rd ed. Sunderland: Sinauer Associates, Inc.

Table 2. A theoretical framework for novice-expert evolutionary reasoning about natural selection and genetic drift.

Level of Expertise	Reasoning Model	Example	Reference
Expert	**Adaptive + Stochastic** Both natural selection and genetic drift collectively explain patterns of evolutionary change.	'…although the initial divergence of Homo from the australopiths may have involved selection, divergence after this time (at least in the facial characters analyzed) could have occurred through random processes alone.' 'This result suggests that both random and to a lesser extent nonrandom processes played an important role in the diversification of this morphologically diverse group; it does not necessarily mean that both played a role across all parts of the group.'	(Ackermann and Cheverud 2004, p. 17951) (Ackermann and Cheverud 2004, p. 17949)
	Adaptive vs. Stochastic Either natural selection or genetic drift leads to evolutionary change.	'QTL data do provide information on the roles of natural selection vs. genetic drift in phenotypic evolution.'	(Orr, 1998, p. 2102)
Emerging Expert	**Adaptive (key concepts only)** Only natural selection	'A mutation may have taken place that allowed a locust to be immune to DDT, this trait was then passed on. These immune locust were the	(Nehm and Ha, 2011) See also this

	explains evolutionary change.	only (ones) that survived and reproduced. Over time, the mutated trait became common of the locust species 'migratoria'.'	paper.
Novice	**Mixed/Synthetic Adaptive (naïve ideas and key concepts)** Naïve ideas and natural selection explain evolutionary change.	'Flightless bird species could have originated from other bird species that can fly because they did not have a specific need for flight. Since they didn't need and/or use their wings for flight, a selective pressure may have worked on them to cause their wings to become flightless.'	(Nehm and Ha, 2011) See also this paper.

Table 3. *Correlations among scores derived from the CINS, ACORNS, and clinical interviews. N = 55. (** p <0.01, * p < 0.05). NAR= Non-adaptive reasoning; KC = Key Concepts; NI = Naive Ideas.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) CINS	1.000							
(2) ACORNS KC	0.426**	1.000						
(3) Interview KC	0.401**	0.566**	1.000					
(4) ACORNS MIS	-0.355**	-0.377**	-0.521**	1.000				
(5) Interview MIS	-0.158	-0.213	-0.354**	0.550**	1.000			
(6) ACORNS NAR	0.017	-0.038	0.046	-0.169	-0.043	1.000		
(7) Mentioning NAR	0.002	-0.089	0.046	-0.158	-0.121	0.861**	1.000	
(8) Scientific NAR	0.079	0.039	0.060	-0.157	-0.068	0.835**	0.851**	1.000

Table 4. *Selected excerpts from students who used non-adaptive reasoning in clinical oral interviews.*

Participant	Non-adaptive terms used	Example use of term	Scientific NAR Score for item excerpted (0-1 scale)	Overall Interview Score (-1, 0 or +1)
A	Bottleneck, founder's effect	'…it could be the bottleneck effect or the founder's effect, like the ancestor species could have had like a nat-, disaster, like a, calamity, so only a few of the, uh, individuals survived and those individuals like maybe had like a dwarf gene, or like a, like a gimp tail I guess. And then eventually it started being more prevalent in the, in the new species that started from there.'	0.5	0
B	Bottleneck	'…Or, um, if there were opossums with shorter tails and something happened to like bottleneck them or something, and it didn't have, uh, an advantage or disadvantage they could just go away randomly.'	0.5	1
C	Founder's effect	(defining the word) 'Um, it's like when, uh, part of the population gets isolated and, uh, um, it's like alleles get uh, a rare allele becomes dominant.'	0	1
D	Founder effect, bottleneck	'…um, I guess that would be a founder effect I think, or a bottleneck. Um, so the species without petals not necessarily that it was favored but because it was in greater variety or greater, uh, it was in greater number than the ones without, that had the larger petals, it eventually, uh, I guess it was eventually derived to a species without petals, in the living species today.'	0	1
E	Genetic drift	'I always forget the definition of genetic drift. Um, it's kinda just like, uh, a swing towards one extreme instead of where it was before, but I guess, so I guess that	0	1

		kind of takes away from what I was going towards, I guess, kinda contradicted myself.'		
F	Bottleneck, founder's effect, genetic drift	'I guess with genetic drift what would happen is, um, like a random grouping of like, let's say out of a hundred only four of these just happen to have, like, this mutation…so over time, like, their genes were the ones that were being passed down randomly, um, into the population, and it just happened to be that as time went on, it was their genes that caused the population, like, not really caused the population but, um, is what the population formed into.'	0	1
G	Gene drift, genetic drift	'you've got your random genetic drift option where there was variation in the presence or absence of teeth in the first species, and then randomly, you know, maybe due to natural disasters, or just certain members got lucky in breeding, you'd have the fixture of the lack, or the presence of teeth.'	1	1
H	Genetic drift	'I think the way genetic drift influence it, is that through time and passing on of traits and genes, that it was affected, oh gosh, I feel like I'm rambling and I'm not really saying anything. (long pause) I can't really think of what I want to say with genetic drift.'	-1	0

Figure 1. **A** *= ACORNS and* **I** *= Interview. (A) Average scores of both majors and advanced majors for the CINS, (B) key concepts in the written ACORNS assessment and Interviews* (*C) naive ideas used in the ACORNS and interviews, and D) Non-adaptive ideas used in the ACORNS and interviews. Though advanced majors show a slight increase in key concepts used and a slight decrease in naive ideas used, this trend is non-significant. However, advanced majors do use significantly more non-adaptive ideas in their evolutionary explanations compared to majors.*

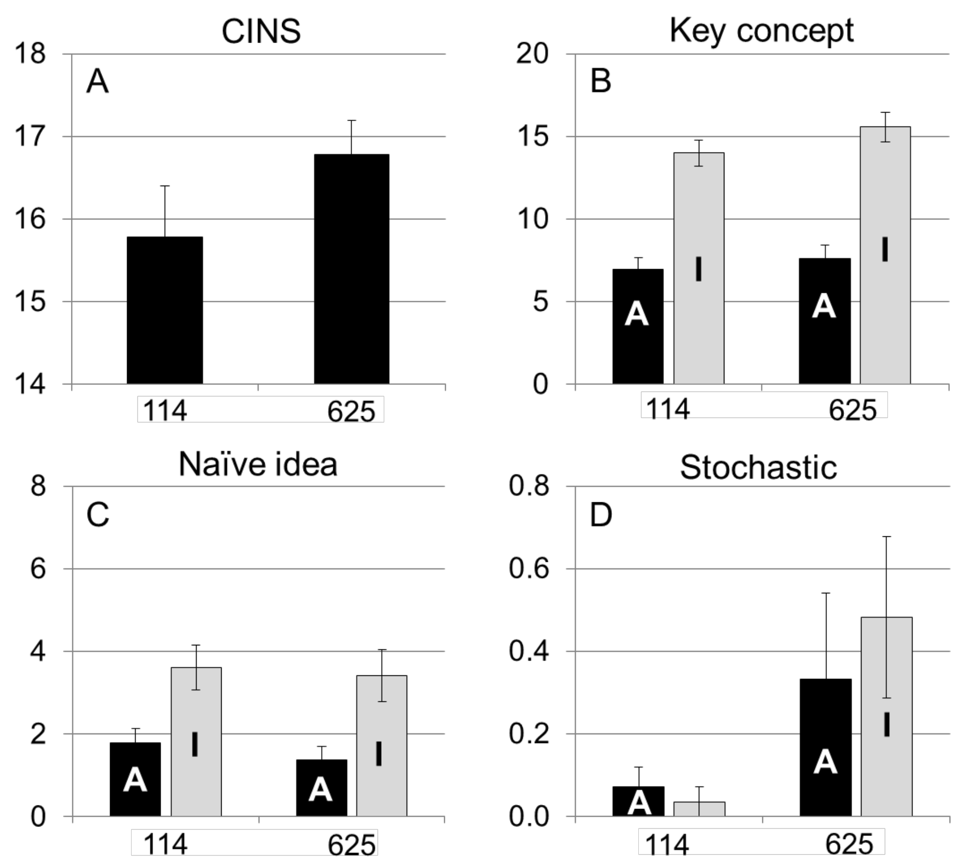

Figure 2. *Diversity and abundance of explanatory elements used in the ACORNS and the clinical interviews. Both key concepts and naive ideas were more abundant in interviews than in the ACORNS. Stochastic reasoning was used more often in the first item than subsequent items.*

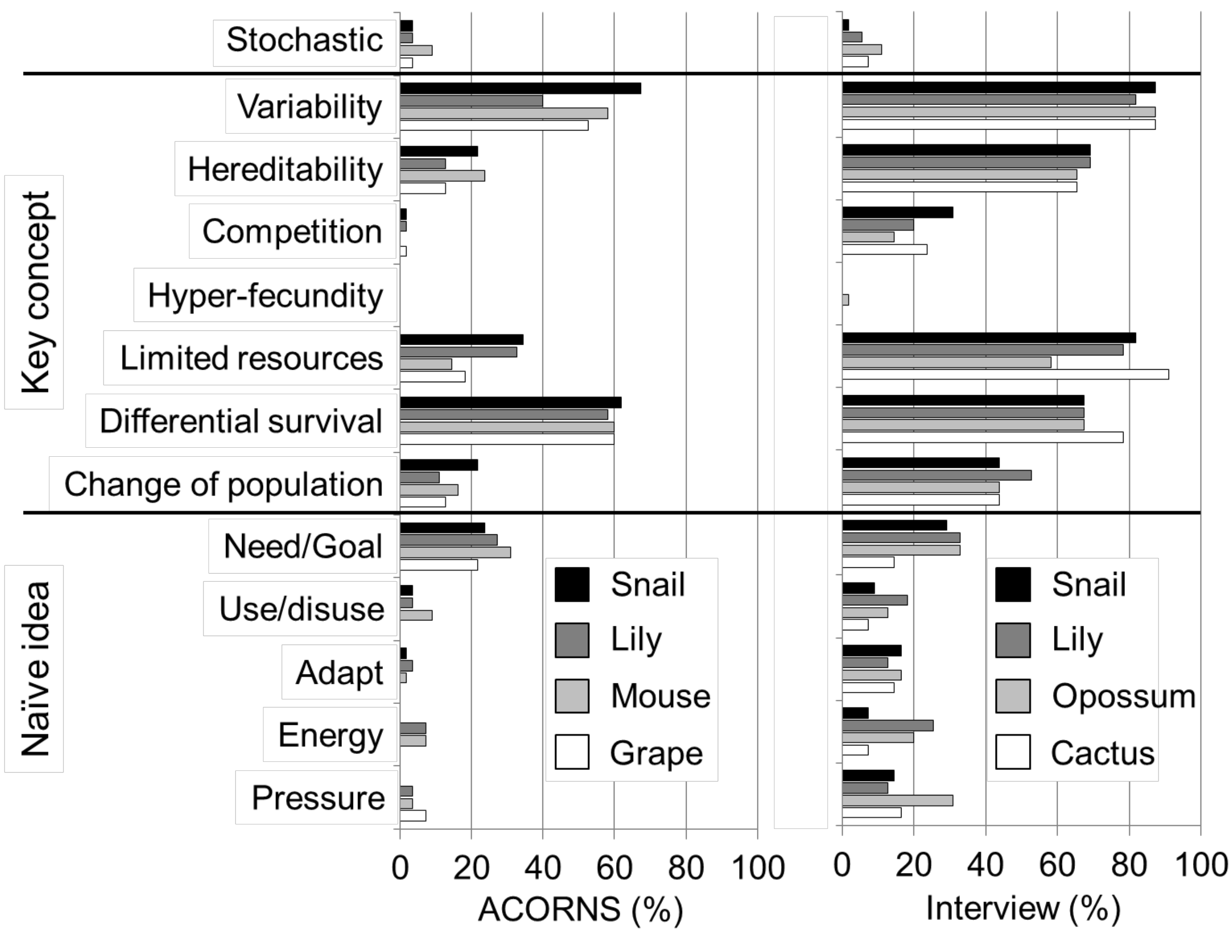

***Figure 3. 114** = majors and **625** = advanced majors. Percentage of students who fall into each category of expertise. The majority of students used a mixed model, though a small portion of these also included NAR into their explanations (shaded gray). A large portion of students also held pure adaptive models and would therefore be considered emerging experts. Only advanced majors reached expert-like levels of reasoning and these students used adaptive vs. stochastic models. No students held selective+stochastic models of evolutionary reasoning.*